CLAIMS

The subject matter claimed is:

1.   A method for predicting antisense activity of an oligonucleotide for down-regulating expression of a selected RNA comprising:

5   (a)   developing an artificial neural network embodied on a computer-readable medium comprising

(i)   constructing a database comprising sequence data of oligonucleotides tested *in vivo* for activity in down-regulating expression of RNAs and activity data corresponding to said sequence data,

10   (ii)   providing an input layer containing a selected number of input nodes, optionally at least one hidden layer comprising a plurality of hidden nodes having full connectivity to said input nodes, and an output layer comprising at least one output node connected to said plurality of hidden nodes, if present, or to said input nodes,

(iii)   mapping sequence motifs of a preselected length found in the sequence

15   data contained in the database, entering counts for each of said sequence motifs in selected input nodes of the input layer, and entering the activity data correlated with said counts of said sequence motifs, and

(iv)   training the artificial neural network having the counts entered in the input layer thereof such that the artificial neural network produces an output in the output layer

20   upon entry of sequence motif counts, wherein said output comprises a measure of predicted activity correlated with sequence motif counts for a test oligonucleotide; and

(b)   mapping sequence motifs of the preselected length present in a nucleotide

38

sequence of a test oligonucleotide complementary to at least a portion of said selected RNA, determining counts of the mapped sequence motifs, and entering the counts of said sequence motifs present in the nucleotide sequence of said test oligonucleotide in the input layer of the artificial neural network; and

5   (c)   obtaining output of the predicted antisense activity of the test oligonucleotide for down-regulating expression of said selected RNA.

2.   The method of claim 1 wherein the sequence data in said database comprise sequence data compiled from published articles wherein each of said published articles reports results obtained with at least ten oligonucleotides and at least one mismatch or scrambled control

10   oligonucleotide.

3.   The method of claim 1 wherein said input layer comprises one input node per sequence motif.

4.   The method of claim 1 wherein said input layer comprises only sequence motifs exhibiting a statistical correlation in their presence to oligonucleotide activity.

15   5.   The method of claim 4 wherein a χ^2 test for significance is performed on the sequence motifs for all oligonucleotide sequences in the database, such sequence motifs are ranked from most to least significant, and the selected number of input nodes corresponds to a selected number of most significant sequence motifs, one input node per most significant

39

sequence motif.

6.   The method of claim 5 wherein said selected number of most significant sequence motifs is about 20 to about 80.

7.   The method of claim 6 wherein said number of most significant sequence motifs is about 40.

8.   The method of claim 1 wherein said at least one hidden layer comprises from about 4 to about 16 hidden nodes.

9.   The method of claim 1 wherein said at least one hidden layer comprises about 4 hidden nodes.

10.   The method of claim 1 wherein said output layer comprises one output node.

11.   The method of claim 1 wherein said training the neural network further comprises using a back-propagation algorithm with a momentum term.

12.   The method of claim 1 wherein said training the neural network further comprises using a back-propagation algorithm without a momentum term.

13.   The method of claim 1 further comprising reporting accuracy of predicted antisense activity by ROC analysis.

14.   The method of claim 1 further comprising assessing generalization of predicted antisense activity by minus 10% cross validation.

15.   The method of claim 1 further comprising assessing generalization of predicted antisense activity by take-one-out cross-validation.

16.   The method of claim 1 further comprising assessing generalization of predicted antisense activity by means of minus-one-RNA cross-validation.

17.   The method of claim 1 wherein said counts of sequence motifs are entered as normalized data.

18.   The method of claim 1 wherein antisense activity of oligonucleotides is entered using a binary threshold function with a cutoff in the range of about 0.01-0.50.

19.   The method of claim 1 wherein discrimination of antisense activity of low-activity oligonucleotides is emphasized and antisense activity of high-activity oligonucleotides is de-emphasized.

20.   The method of claim 1 further comprising combining the predicted antisense

activity of the artificial neural network with a predicted antisense activity of at least one other

artificial neural network.

21.   The method of claim 1 further comprising combining the predicted antisense

5   activity of the artificial neural network with an estimator of free-energy change associated with

oligonucleotide-RNA duplex creation.

22.   A method of making an artificial neural network, embodied on a computer-

readable medium, for predicting antisense activity of oligonucleotides for down-regulating

expression of a selected RNA comprising:

(a)   constructing a database comprising sequence data of oligonucleotides tested *in*

10   *vivo* for activity in down-regulating expression of RNAs and activity data corresponding to said

sequence data;

(b)   providing an input layer containing a selected number of input nodes, optionally at

least one hidden layer comprising a plurality of hidden nodes having full connectivity to said

15   input nodes, and an output layer comprising at least one output node connected to said plurality

of hidden nodes, if present, or to said input nodes;

(c)   mapping sequence motifs of a preselected length found in the sequence data

contained in the database, entering counts for each of said sequence motifs in selected input

nodes of the input layer, and entering the activity data correlated with said counts of said

20   sequence motifs; and

42

(d)   training the artificial neural network having the counts entered in the input layer

thereof such that the artificial neural network produces an output in the output layer upon entry of

sequence motif counts, wherein said output comprises a measure of predicted activity correlated

with sequence motif counts for a test oligonucleotide.

5   23.   The method of claim 22 wherein the sequence data in said database comprise

sequence data compiled from published articles wherein each of said published articles reports

results obtained with at least ten oligonucleotides and at least one mismatch or scrambled control

oligonucleotide.

24.   The method of claim 22 wherein said input layer comprises one input node per

10   sequence motif.

25.   The method of claim 22 wherein said input layer comprises only sequence motifs

exhibiting a statistical correlation in their presence to oligonucleotide activity.

26.   The method of claim 25 wherein a χ^2 test for significance is performed on the

sequence motifs for all oligonucleotide sequences in the database, such sequence motifs are

15   ranked from most to least significant, and the selected number of input nodes corresponds to a

selected number of most significant sequence motifs, one input node per most significant

sequence motif.

43

27.   The method of claim 26 wherein said selected number of most significant sequence motifs is about 20 to about 80.

28.   The method of claim 27 wherein said number of most significant sequence motifs is about 40.

29.   The method of claim 22 wherein said at least one hidden layer comprises from about 4 to about 16 hidden nodes.

30.   The method of claim 22 wherein said at least one hidden layer comprises 4 hidden nodes.

31.   The method of claim 22 wherein said output layer comprises one output node.

32.   The method of claim 22 wherein said training the neural network further comprises using a back-propagation algorithm with a momentum term.

33.   The method of claim 22 wherein said training the neural network further comprises using a back-propagation algorithm without a momentum term.

34.   The method of claim 22 further comprising reporting accuracy of predicted antisense activity by ROC analysis.

44

35. The method of claim 22 further comprising assessing generalization of predicted antisense activity by minus 10% cross validation.

36. The method of claim 22 further comprising assessing generalization of predicted antisense activity by take-one-out cross-validation.

5

37. The method of claim 22 further comprising assessing generalization of predicted antisense activity by means of minus-one-RNA cross-validation.

38. The method of claim 22 wherein said counts of sequence motifs are entered as normalized data.

39. The method of claim 22 wherein antisense activity of oligonucleotides is entered

10 using a binary threshold function with a cutoff in the range of about 0.01-0.50.

40. The method of claim 22 wherein discrimination of antisense activity of low-activity oligonucleotides is emphasized and antisense activity of high-activity oligonucleotides is de-emphasized.

41. The method of claim 22 further comprising combining the predicted antisense

15 activity of the artificial neural network with a predicted antisense activity of at least one other artificial neural network.

45

42.   The method of claim 22 further comprising combining the predicted antisense

activity of the artificial neural network with an estimator of free-energy change associated with

oligonucleotide-RNA duplex creation.

43.   An artificial neural network embodied on a computer-readable medium made by

5   the method of claim 22.

44.   An artificial neural network embodied on a computer-readable medium

comprising:

(a)   an input layer containing a selected number of input nodes;

(b)   optionally at least one hidden layer comprising a plurality of hidden nodes having

10   full connectivity to said input nodes; and

(c)   an output layer comprising at least one output node connected to said plurality of

hidden nodes, if present, or to said input nodes;

wherein sequence motifs of a preselected length found in a database comprising (i)

sequence data of oligonucleotides tested *in vivo* for activity in down-regulating expression of

15   RNAs and (ii) activity data corresponding to said sequence data are mapped and counts for each

of said mapped sequence motifs are entered in selected input nodes of the input layer, and the

activity data correlated with said counts of said sequence motifs are also entered in said selected

input nodes of the input layer, and then the artificial neural network is trained such that the

artificial neural network produces an output in the output layer upon entry of sequence motif

20   counts, wherein said output comprises a measure of predicted activity correlated with sequence

46

motif counts for a test oligonucleotide.

45.   The artificial neural network of claim 44 wherein the sequence data in said

database comprise sequence data compiled from published articles wherein each of said

published articles reports results obtained with at least ten oligonucleotides and at least one

5   mismatch or scrambled control oligonucleotide.

46.   The artificial neural network of claim 44 wherein said input layer comprises one

input node per sequence motif.

47.   The artificial neural network of claim 44 wherein said input layer comprises only

sequence motifs exhibiting a statistical correlation in their presence to oligonucleotide activity.

48.   The artificial neural network of claim 47 wherein a χ^2 test for significance is

performed on the sequence motifs for all oligonucleotide sequences in the database, such

sequence motifs are ranked from most to least significant, and the selected number of input nodes

corresponds to a selected number of most significant sequence motifs, one input node per most

significant sequence motif.

15   49.   The artificial neural network of claim 48 wherein said selected number of most

significant sequence motifs in about 20 to about 80.

50.   The artificial neural network of claim 49 wherein said number of most significant sequence motifs is about 40.

51.   The artificial neural network of claim 44 wherein said at least one hidden layer comprises from about 4 to about 16 hidden nodes.

5   52.   The artificial neural network of claim 44 wherein said at least one hidden layer comprises 4 hidden nodes.

53.   The artificial neural network of claim 44 wherein said output layer comprises one output node.

54.   The artificial neural network of claim 44 wherein said artificial neural network is

10   trained using a back-propagation algorithm with a momentum term.

55.   The artificial neural network of claim 44 wherein said artificial neural network is trained using a back-propagation algorithm without a momentum term.

56.   The artificial neural network of claim 44 wherein accuracy of predicted antisense activity is reported by ROC analysis.

15   57.   The artificial neural network of claim 44 wherein generalization of predicted

48

antisense activity is assessed by minus 10% cross validation.

58.   The artificial neural network of claim 44 wherein generalization of predicted

antisense activity is assessed by take-one-out cross-validation.

59.   The artificial neural network of claim 44 generalization of predicted antisense

5   activity is assessed by means of minus-one-RNA cross-validation.

60.   The artificial neural network of claim 44 wherein said counts of sequence motifs

are entered as normalized data.

61.   The artificial neural network of claim 44 wherein antisense activity of

oligonucleotides is entered using a binary threshold function with a cutoff in the range of about

10   0.01-0.50.

62.   The artificial neural network of claim 44 wherein discrimination of antisense

activity of low-activity oligonucleotides is emphasized and antisense activity of high-activity

oligonucleotides is de-emphasized.

63.   The artificial neural network of claim 44 wherein the predicted antisense activity

15   is combined with a predicted antisense activity of at least one other artificial neural network.

64.   The artificial neural network of claim 44 wherein the predicted antisense activity

is combined with an estimator of free-energy change associated with oligonucleotide-RNA

duplex creation.